Reducing Autocorrelation Times in HMC Simulation

S. Yamamoto[1] P. A. Boyle[1] T. Izubuchi[1] L. Jin[2] C. Lehner[3] N. Matsumoto[4]

[1]Brookhaven National Laboratory

[2]Department of Physics, University of Connecticut

[3]Fakultät für Physik, Universität Regensburg

[4]Department of Physics, University of Boston

Brookhaven Forum 2025

Table of Contents

Table of Contents

Lattice QCD

In lattice QCD, $\langle O \rangle$ is evaluated non-perturbatively.

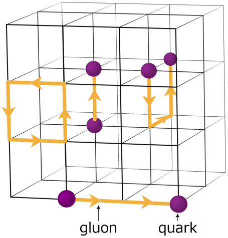
- Wick-rotated

$$\langle O \rangle = \frac{1}{Z} \int \mathcal{D}\phi O(\phi) e^{-S_{\mathrm{E}}}$$

- Then, we use $e^{-S_{\mathrm{E}}}$ as a Monte-Carlo weight

To make the computation doable on a computer,

- We discretize the space-time into the lattice with its spacing a and dimension $L^3 \times T$
- Fermion fields, ψ, live on a lattice site, and gauge fields are replaced by a link, U_μ, connecting the adjacent points.



[1]Picture courtesy of Simone Bacchio

Lattice QCD

- Configurations are sampled using importance sampling method

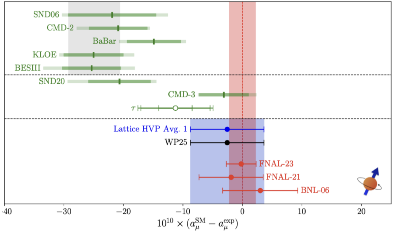
$$\langle O \rangle = \lim_{N\to\infty} \frac{1}{N} \sum_{i}^{N} O_i = \frac{1}{N} \sum_{i}^{N} O_i + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right)$$

- Sample is generated using Markov Chain Monte Carlo:
  - Markov Property: The probability of a event happening at a given step depends only on the outcome of the previous step and not on the history of events
  - Rosenbluth's, Teller's and Metropolis developed a Markov chain Monte Carlo sampler, known as Metropolis algorithm at Los Alomos under precursor to DOE.
  - Broadly used in physics, math, statistics, computer science and machine learning.
- Increasing N reduces statistical error in lattice calculation

Muon g-2

2025 Theory Initiative White Paper

Contribution	Section	Equation	Value $\times 10^{11}$
Experiment (E989)		Eq. (9.5)	116 592 059(22)
HVP LO (lattice)	Sec. 3.6.1	Eq. (3.37)	7132(61)
HVP LO (e^+e^-, τ)	Sec. 2	Table 5	Estimates not prov
HVP NLO (e^+e^-)	Sec. 2.9	Eq. (2.47)	−99.6(1.3)
HVP NNLO (e^+e^-)	Sec. 2.9	Eq. (2.48)	12.4(1)
HLbL (phenomenology)	Sec. 5.10	Eq. (5.69)	103.3(8.8)
HLbL NLO (phenomenology)	Sec. 5.10	Eq. (5.70)	2.6(6)
HLbL (lattice)	Sec. 6.2.8	Eq. (6.34)	122.5(9.0)
HLbL (phenomenology + lattice)	Sec. 9	Eq. (9.2)	112.6(9.6)
QED	Sec. 7.5	Eq. (7.27)	116 584 718.8(2)
EW	Sec. 8	Eq. (8.12)	154.4(4)
HVP LO (lattice) + HVP N(N)LO (e^+e^-)	Sec. 9	Eq. (9.1)	7045(61)
HLbL (phenomenology + lattice + NLO)	Sec. 9	Eq. (9.3)	115.5(9.9)
Total SM Value	Sec. 9	Eq. (9.4)	116 592 033(62)
Difference: $\Delta a_\mu \equiv a_\mu^{\text{exp}} - a_\mu^{\text{SM}}$	Sec. 9	Eq. (9.6)	26(66)



$$a_\mu(\text{Run-4/5/6}) = 116\,592\,0710(162) \times 10^{-12} \quad (139\,\text{ppb})$$

$$a_\mu(\text{Run-1-6}) = 116\,592\,0705(148) \times 10^{-12} \quad (127\,\text{ppb}).$$

$$a_\mu(\text{WP25}) = 1165920330(620) \times 10^{-12} \quad (532\,\text{ppb})$$

- FNAL published the result in 2025 that shows reduced error to 127 ppb without a shift of the central value
- Lattice calculation is now the preffered method to compute LO Hadronic contribution

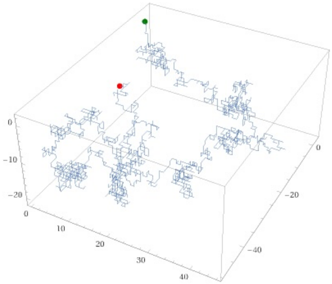
HMC

- HMC = Hybrid Monte Carlo = **Molecular Dynamics** (MD) with Momentum Refreshment + **Monte Carlo** (MC)
- Insert a constant factor to Z:

$$Z = \int \mathcal{D}U\, e^{-S(U)} \propto \int \mathcal{D}U \mathcal{D}P\, e^{-P^2/2} e^{-S(U)} = \int \mathcal{D}U \mathcal{D}P\, e^{-H}$$

- $H = P^2/2 + S(U)$
- MD Part
  - Momentum Refreshment: Draw P according to $e^{-P^2/2}$
  - MD Evolution: Solve the equation of motion for H in the (U, P)-space
- Accept-Reject Step:
  - Accept U' as the new link in the MC chain with a probability or else $U_{n+1} = U$
  - **Ensures the resulting distribution is the target distribution**
  - **Remove a source of systematic uncertainty, e.g., from numerical integration of MD evolution**

Autocorrelation



Autocorrelation Time

$$\sigma^2(\bar{A}) = \frac{\sigma^2(A)}{N}\left[1 + 2\sum_{t=1}^{N-1}\left(1 - \frac{t}{N}\right)C(t)\right] = \frac{\sigma^2(A)}{N}\tau_{int} = \frac{\sigma^2(A)}{N/\tau_{int}}$$

- Variance of the mean is reduced not by the factor N but by N/τ_{int}
- Autocorrelation reduces the effective size of the sample by τ_{int}

Critical Slowing Down

- Autocorrelation times typically increase approximately like a^{-2} as $a \to 0$ [1]
  - a is the lattice spacing
- In the free field analysis, τ_{int} is larger for observables with longer correlation length [2]
  - Analogy: Simple Harmonic Oscillator: $H = \sum \tilde{\pi}_p^2 + \omega_p^2 \tilde{\phi}_p$ where $\omega_p^2 = m^2 + p^2$
- Also, as $a \to 0$, MD trajectory is observed to be trapped within a topological sector in a practical simulation, a phenomenon known as topological freezing

[1] [Luscher, 2010]

[2] [Kennedy and Pendleton, 2001]

Field-Transformation HMC

With $U = \mathcal{F}_t(V)$,

$$Z = \int \mathcal{D}U e^{-S(U)} = \int \mathcal{D}V \mathrm{Det}[\mathcal{F}_*(V)]\, e^{-S(\mathcal{F}(V))} = \int \mathcal{D}V\, e^{-S_{FT}(V)}$$

$$S_{FT} = S(\mathcal{F}_t(V)) - \ln \mathrm{Det}\mathcal{F}_*(V).$$

- originally proposed by Luscher for continuous flow [Lüscher, 2010]
- perfect trivialization: $S_{FT} = 0$

In our study,

- approximate the trivializing map by the Wilson flow
- discretize the transformation with step of size ρ
- The number of integration steps for the discretized trivializing map is set to 1

Table of Contents

Statistics

ρ	0.0	0.1	0.112	0.124
$\delta\tau_G = 1/48$	233	230	188	230
$\delta\tau_G = 1/96$	401	232	229	229
$\delta\tau_G = 1/144$	-	230	-	-

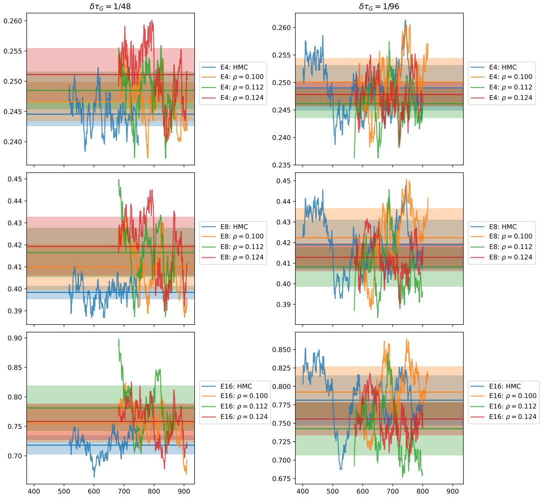
Table: The number of configurations for each ensemble after thermalization

Machine

- Simulation is carried out on Frontier and Andes at Oak Ridge National Laboratory
- Analysis of data is performed on Aurora at Argonne Natiobal Laboratory

Table of Contents

Observable: Wilson flowed energies



- Comparison of Wilson flowed energy with different ρ values for different flow time (raw) and $\delta\tau_G = 1/48, 1/96$ (column)

Table of Contents

Autocorrelation for Local Quantities

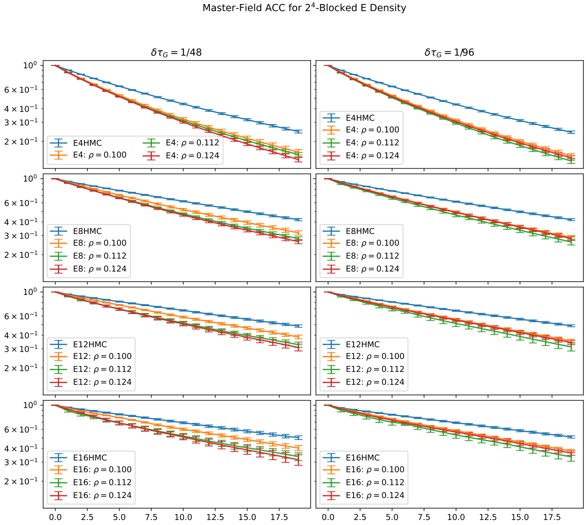


Figure: Autocorrelation based on Master-Field technique

Autocorrelation Times

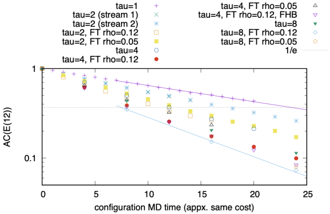
The ratios of $\tau_{\text{exp}}(\rho = 0.0, \delta\tau_G = 1/48)$ for HMC to τ_{exp} with other HMC parameters:

ρ	$\tau_W = 4$	$\tau_W = 16$
0.100	1.275	1.2832
0.112	1.313	1.4487
0.124	1.408	1.5736

Table: Fixed $\delta\tau_G = 1/48$, varied Wilson flow time τ_W

Longer Trajectory Length Decreases Autocorrelation

Christoph Lehner (2024):



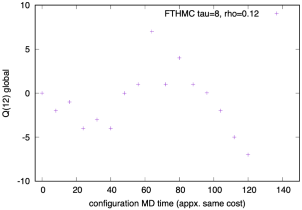
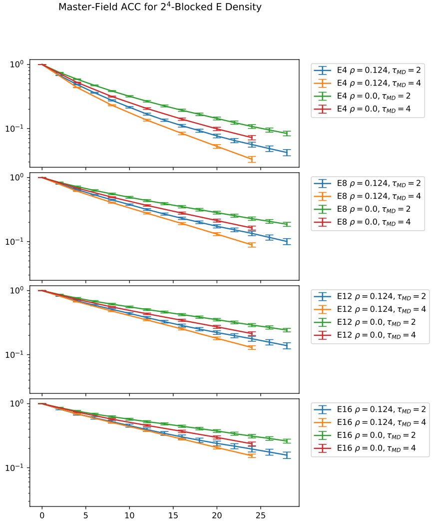


Figure: Autocorrelation of Wilson energy with $\tau_W = 12$ vs. MC time in MD units, measured on $62^3 \times 96$ lattice (left) and topological charge Q vs. MC time, measured on $128^3 \times 288$ lattice (right)

- GPT implementation is used for computation
- Q is measured on the lattice at the physical point $a^{-1} \approx 3.5$ GeV with $2 + 1f$ DWF
- The effec of longer trajecory length is additive

Longer Trajectory Length Decreases Autocorrelation



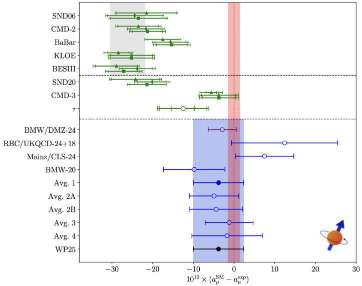
τ_{MD}	$\tau_W = 12$
1	1.574(10)
2	1.817(2)
4	2.139(2)

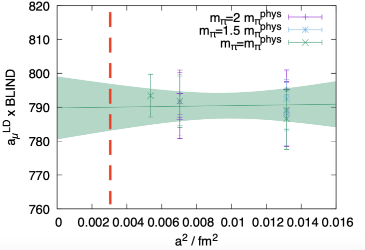
Table: The ratio of autocorrealtion time with traditional HMC to the ones with different trajectory length τ_{MD} and fixed $\delta\tau_G = 1/48$, $\rho = 0.124$, $\tau_W = 12$

Table of Contents

Summary

- FTHMC reduces autocorrelation times around 1.5x compared to HMC
- Longer trajectory length + field transformation reduced autocorrelation time around 3.5x
- Enabled simulation at finer lattice spacings, huge volume (3.5 Gev, $128^3 \times 288$ for Iwasaki gauge action and 2+1 DWF)
- $\Rightarrow$ continue with gauge + fermion action: big impact on physics program of RBC-UKQCD





Outlook

- FTHMC appears to mitigate topological freezing
- currently exploring what causes enhancement of frequency of transitions to different topological sectors

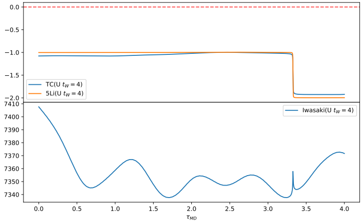


Figure: The plot of flowed topological charge vs. τ_{MD} (top) and flowed gauge action vs. τ_{MD} (bottom)

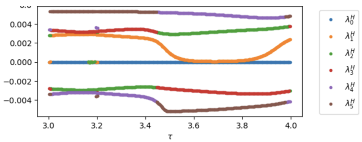


Figure: near-zero eigenvalues of Dirac operator vs. τ_{MD}

Thank you!

Bruno, M., Cè, M., Francis, A., Fritzsch, P., Green, J. R., Hansen, M. T., and Rago, A. (2023).
Exploiting stochastic locality in lattice QCD: hadronic observables and their uncertainties.
JHEP, 11:167.

Kennedy, A. D. and Pendleton, B. (2001).
Cost of the generalized hybrid Monte Carlo algorithm for free field theory.
Nucl. Phys. B, 607:456–510.

Luscher, M. (2010).
Computational Strategies in Lattice QCD.
In *Les Houches Summer School: Session 93: Modern perspectives in lattice QCD: Quantum field theory and high performance computing*, pages 331–399.

Lüscher, M. (2010).

Summary and Outlook

- Master-Field technique allows us to measure autocorrelation coefficients based on a small number of configurations
- Generate ensemble with different parameters (*beta*, the number of trivializing steps, etc...) for tuning
- FTHMC showed potential to reduce autocorelation time for topologial charge
- However, there are a number of parameters for FTHMC to tune for optimal performance
- Better understanding of why and how FTHMC is effective is needed

Acceptance Rates

	HMC	$\rho = 0.1$	$\rho = 0.112$	$\rho = 0.124$
$\delta\tau_G = 1/48$	0.929(6)	0.944(5)	0.935(6)	0.924(6)
$\delta\tau_G = 1/96$	-	0.956(4)	0.944(5)	0.94(5)

Table: $\langle P_{\mathrm{acc}} \rangle$ for runs with and without FT.

Markov Chains

Markov chain: $\{X_n\}_{n=0}^{N}$

- A sequence of random variables with Markov Property:

$$\Pr\{X_n = x_n | X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \cdots, X_0 = x_0\}$$
$$= \Pr\{X_n = x_n | X_{n-1} = x_{n-1}\} = P_n^{x_{n-1}}(x_n) = \frac{P_{X_n \cap X_{n-1}}(x_n, x_{n-1})}{P_{X_{n-1}}(x_{n-1})}$$

- In words, the probability of a event happening depends only on the outcome of the last outcome and not on the history of events
- Here, $P_n^{x_{n-1}} : E \to [0,1]$
- P_0 is an initial distribution corresponding to an independent random variable
- Transition Probability, $T_n : E \times E \to [0,1]$ via $T_n(x_n, x_{n-1}) = P_n^{x_{n-1}}(x_n)$

Markov Chains

In QCD, we consider time-homogeneous Markov Chain, i.e.,

- $\Pr\{X_{n+1} = x_{n+1} | X_n = x_n\} = \Pr\{X_n = x_n | X_{n_1} = x_{n-1}\} \quad \therefore T_n = T$
- Also, $P_n(x_n) = T^n P_0(x_n)$

We assume

- our Markov chains is irreducible
- all states are aperiodic, $(\forall s \in E \, \forall N \in \mathbb{N}) \; T^N(s \to s) \neq 0$ and positive recurrent $E[\tau_{\mathrm{recurrence}}] < \infty$

Then, [Rothe, 2012]

- There exists a stationary distribution π, and it is unique
- if the initial distribution is π, it is (wide-sense) stationary (WSS)
- if we further have $E[\tau^2_{\mathrm{recurrence}}] < \infty$,

$$\langle O \rangle = (1/N) \sum_{i=1}^{N} O(x_i) + \mathcal{O}(1/\sqrt{N})$$

How do we find such T with a desired distribution π?

The Acceptance-Rejection Method

$T(i \to j) = T_0(i \to j) P_{acc}(i,j) + \delta_{ij} \sum_k T_0(i \to j)[1 - P_{\mathrm{acc}}(i,j)]$

- T_0: a transition matrix with micro-reversibility, $T_0(s \to s') = T_0(s' \to s)$
- $P_{acc}(i,j) = \min\{1, \pi(i)/\pi(j)\}$

Then, the stationary distribution of T is the target distribution π.

Also, it satisfies detailed balance condition:

$$P(i) T(i \to j) = P(j) T(j \to i)$$

$SU(3)$

For QCD,

- $\pi(U) = \frac{1}{Z} e^{-S_G(U)}$
- $Z = \int \mathcal{D}U e^{-S_G(U)}$

Requirement: [Luscher, 2010]

- $T(U \to U') \geq 0$ for all U, U' and $\int \mathcal{D}U' \; T(U \to U') = 1$ for all U.
- $\int \mathcal{D}U \; \pi(U) T(U \to U') = \pi(U')$ for all U'
- $\forall V \; \exists \mathcal{N}_V$, where $\mathcal{N}_V$ is an open neighborhood of V in the space of gauge configurations, s.t.

$$\forall U, U' \in \mathcal{N}_V \; \exists \varepsilon > 0 \text{ s.t. } \; T(U \to U') \geq \varepsilon$$

HMC

- HMC = Hybrid Monte Carlo = Molecular Dynamics (MD) with Momentum Refreshment + Monte Carlo (MC)
- In MD, a partition function of classical statistical system is approximated by trajectories of the canonical Hamilton system by using ergodicity
- To use the technique of MD,

$$Z = \int \mathcal{D}U\, e^{-S(U)} \propto \int \mathcal{D}U\mathcal{D}P\, e^{-P^2/2} e^{-S(U)} = \int \mathcal{D}U\mathcal{D}P\, e^{-H}$$

- $H = P^2/2 + S(U)$

HMC Steps

1. A momentum field P is generated randomly with probability density proportional to $e^{-P^2/2}$
2. The Hamilton equations are integrated from time $t = 0$ to some later time τ with the initial fields of P and U to obtain a new field U'
3. Apply the Acceptance-Reject step to decide whether to set U_τ to U' or keep U, i.e., $U_\tau = U$
4. Repeat

The above steps correspond to

$$T_0(U \to U') = \frac{1}{\mathcal{Z}_P} \int \mathcal{D}P\, e^{-P^2/2} \prod_{x,\mu} \delta(U'(x,\mu)U(x,\mu))$$

HMC Steps

1. A momentum field P is generated randomly with probability density proportional to $e^{-P^2/2}$
2. The Hamilton equations are integrated from time $t = 0$ to some later time τ with the initial fields of P and U to obtain a new field U'
3. Accept U' and set $U_n = U_n$ with probability $P_{acc} = \min\{1, e^{S_G(U) - S_G(U')}\}$. Otherwise, keep U, i.e., $U_n = U$
4. Repeat

Step (1) and (2) correspond to

$$T_0(U \to U') = \frac{1}{\mathcal{Z}_P} \int \mathcal{D}P \, e^{-P^2/2} \prod_{x,\mu} \delta(U'(x,\mu)U(x,\mu))$$

Numerical Integration

Elementary updates for P and U

$$I_P(\varepsilon):(P,U) \to (P-\varepsilon F, U)$$
$$I_U(\varepsilon):(P,U) \to (P, e^{\varepsilon P}U)$$

Leap-frog integrator:

$$\mathcal{J}(\varepsilon, N) = \{I_P(\varepsilon/2) I_U(\varepsilon) I_P(\varepsilon/2)\}^N$$

where $\varepsilon = \tau/N$

Run Parameters

Lattice Parameters:

- on a lattice of size 32^4
- $\beta = 2.37$
- with $2+1$ Domain-Wall fermions of mass $m_l = 0.0047, m_s = 0.0186$

HMC Parameters:

- different ρ values: 0.1, 0.112, 0.124
- different gauge step sizes $\delta\tau_G = 1/48, 1/96$
- different fermion step sizes $\delta\tau_F = 1/24, 1/16, 1/12, 1/8$

In the following, we focus on the runs with different flow parameters and $\delta\tau_G$

Master-Field Technique

- Instead of ACC of the volume average $\langle\!\langle A \rangle\!\rangle = (1/V)\sum_x A(x)$, consider ACC of local observable $A(x)$
- Idea: $\langle\!\langle A(x) \rangle\!\rangle = \langle A(x) \rangle + \mathcal{O}(V^{-1/2})$
- Approximate $\Gamma'_x(t)$ by $\langle\!\langle \Gamma'(t) \rangle\!\rangle$ [Lüscher, 2018]
    - Autocovariance of $A'(x)$ at x: $\Gamma'_x(t)$
    - The volume average is subtracted: $A'(x) = A(x) - \langle\!\langle A \rangle\!\rangle$

Master-Field Technique

- Idea: $\langle\langle A(x) \rangle\rangle = \langle A(x) \rangle + \mathcal{O}(V^{-1/2})$
- Approximate $\Gamma'_x(t)$ by $\langle\langle \Gamma'(t) \rangle\rangle$ [Lüscher, 2018]
- Also, $\mathcal{O}^i_t(x) \to \bar{\mathcal{O}}_t(x) \equiv \frac{1}{T-t} \sum_{i=1}^{T-t} \mathcal{O}^i_t(x)$
- Finally, $\rho(t) = \langle\langle \Gamma'(t) \rangle\rangle / \langle\langle \Gamma'(0) \rangle\rangle$

Master-Field Technique

Estimators

- Replace ensemble average by the average over MC chain: $\langle A(x) \rangle \to \bar{a}(x) = (1/T) \sum_i a_i(x)$
- Treat $\bar{\Gamma}'_x(t)$ at different x as correlated but distinct measurements of autocovariance of local observable $A'(x)$
- Consider: $\langle\!\langle \bar{\Gamma} \rangle\!\rangle(t) = (1/V) \sum_x \bar{\Gamma}_x(t)$
- need to take into account lattice-correlation when estimating the error

Volume Autocorrelation

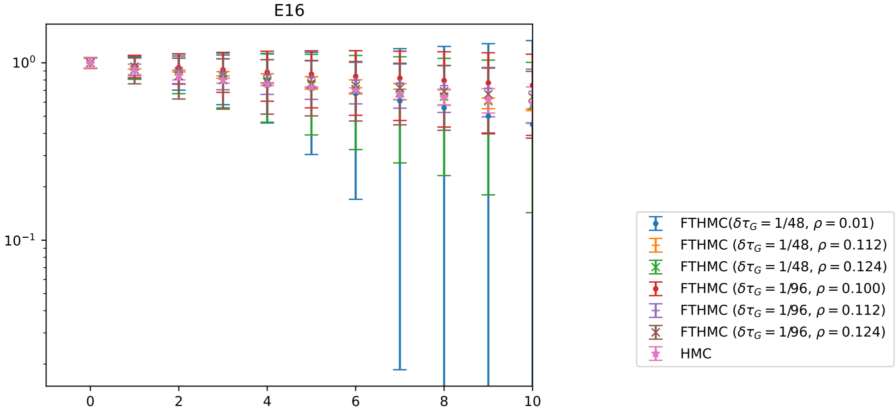


Figure: Autocorrelation coefficient (ACC) as a function of t for Wilson-flowed energy E16.

Autocorrelation for Local Quantities

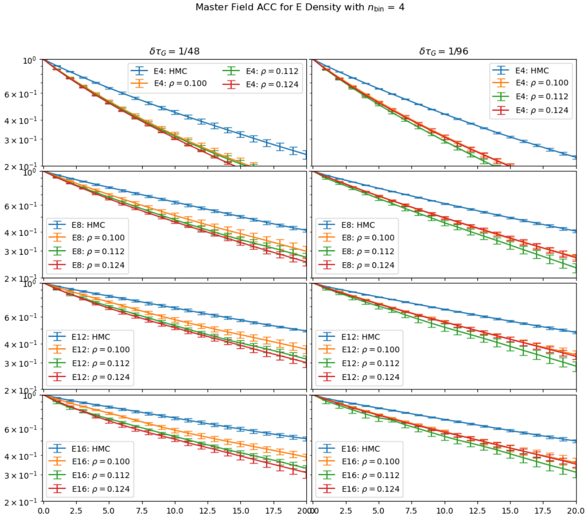


Figure: Autocorrelation based on binning method

Error via Master-Field Approach

- Need: $\mathrm{Cov}[\langle\!\langle \bar{\mathcal{O}}_s \rangle\!\rangle, \langle\!\langle \bar{\mathcal{O}}_t \rangle\!\rangle] \equiv \langle [\langle\!\langle \bar{\mathcal{O}}_s \rangle\!\rangle - \langle \mathcal{O}_s \rangle][\langle\!\langle \bar{\mathcal{O}}_t \rangle\!\rangle - \langle \mathcal{O}_t \rangle] \rangle =$ $\frac{1}{V}\sum_y \langle [\bar{\mathcal{O}}_s(y) - \langle \mathcal{O}_s \rangle][\bar{\mathcal{O}}_t(0) - \langle \mathcal{O}_t \rangle] \rangle \equiv \frac{1}{V}\sum_y C_{st}(y)$ [Bruno et al., 2023]
- Approximate $C_{st}(y)$ by

$$\langle\!\langle \mathcal{C}_{st}(y) \rangle\!\rangle = \frac{1}{V}\sum_x \delta\bar{\mathcal{O}}_s(x+y)\delta\bar{\mathcal{O}}_t(x), \ \delta\bar{\mathcal{O}}_t(x) \equiv \bar{\mathcal{O}}_t(x) - \langle\!\langle \bar{\mathcal{O}}_t \rangle\!\rangle$$

- Define $C_{st}(|y| \le R) \equiv \sum_{|y|\le R} C_{st}(y)$
- Determine the value of R s.t. $C_{st}(|y| \le R)$ saturates
- Truncate the sum in $\mathrm{Cov}[\langle\!\langle \bar{\mathcal{O}}_s \rangle\!\rangle, \langle\!\langle \bar{\mathcal{O}}_t \rangle\!\rangle]$ beyond R_{sat}

$$\mathrm{Var}[\rho(t))] = (\rho(t))^2 \left(\frac{\mathrm{Var}[\langle\!\langle \bar{\Gamma}(t) \rangle\!\rangle]}{\langle\!\langle \bar{\Gamma}(t) \rangle\!\rangle^2} + \frac{\mathrm{Var}[\langle\!\langle \bar{\Gamma}(0) \rangle\!\rangle]}{\langle\!\langle \bar{\Gamma}(0) \rangle\!\rangle^2} - 2\frac{\mathrm{Cov}[\langle\!\langle \bar{\Gamma}(t) \rangle\!\rangle, \langle\!\langle \bar{\Gamma}(0) \rangle\!\rangle]}{\langle\!\langle \bar{\Gamma}(t) \rangle\!\rangle \langle\!\langle \bar{\Gamma}(0) \rangle\!\rangle} \right)$$

Error via Master-Field Approach

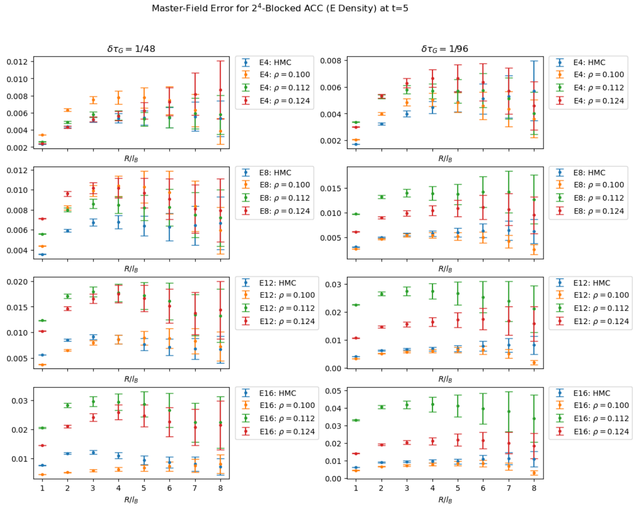


Figure: R: Summation Radius, b: block size